

NORMALIZAÇÃO

Lílian Simão Oliveira

Normalização

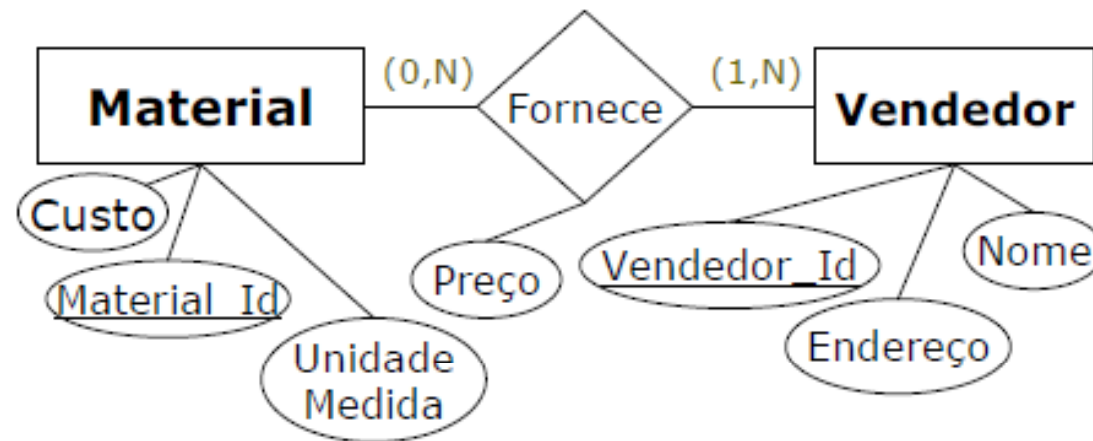


- Normalização de dados : decomposição de esquemas para minimizar redundância e evitar anomalias de atualização, inserção e remoção.
- Projeto conceitual bem feito resulta naturalmente em esquemas normalizados.
- Motivação:
 - ▣ Melhora o projeto lógico evitando duplicações desnecessárias de dados
 - ▣ Produz relações menores mais bem estruturadas

Dependência Funcional

- E uma restrição entre dois atributos ou dois conjuntos de atributos
- Advêm do conhecimento da semântica dos dados armazenados
- Para qualquer relação R ,
 - ▣ um atributo B é funcionalmente dependente de um atributo A se, para toda instância de R , um valor de A determina um único valor de B
 - ▣ ou
$$A \rightarrow B$$
$$\{A, B\} \rightarrow \{C, D\} \approx AB \rightarrow CD$$
- Determinante: atributo do lado esquerdo da flecha

Dependência Funcional



Material

<u>Material_Id</u>	Custo	Unidade_de_Medida
--------------------	-------	-------------------

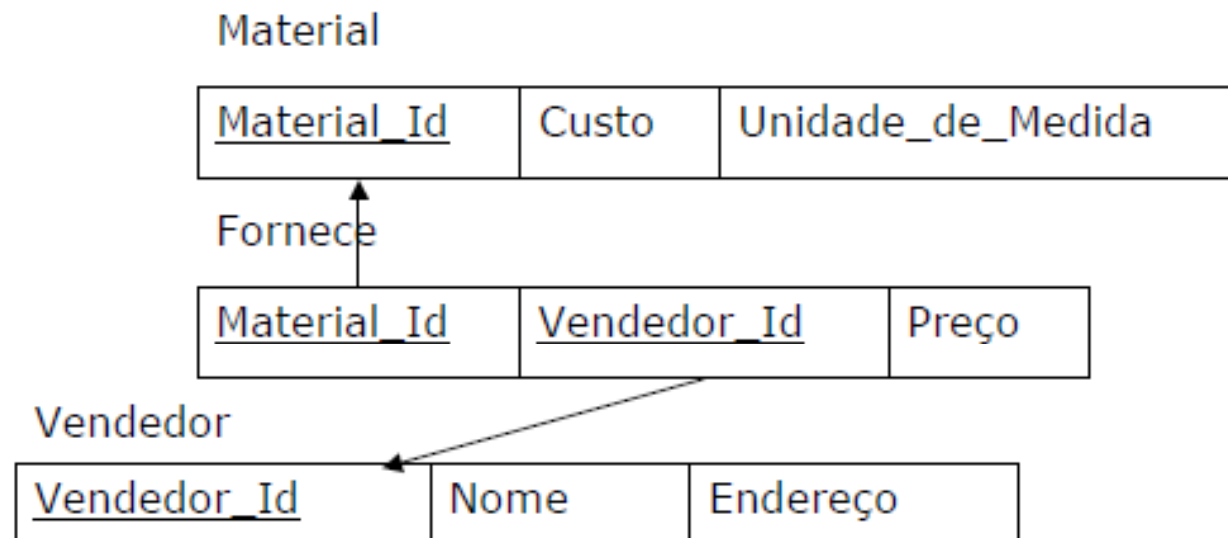
Fornece

<u>Material_Id</u>	<u>Vendedor_Id</u>	Preço
--------------------	--------------------	-------

Vendedor

<u>Vendedor_Id</u>	Nome	Endereço
--------------------	------	----------

Dependência Funcional



- $\text{Material_Id} \rightarrow \{\text{Custo}, \text{Unidade_de_Medida}\}$
- $\text{Vendedor_Id} \rightarrow \{\text{Nome}, \text{Endereço}\}$
- $\{\text{Material_Id}, \text{Vendedor_Id}\} \rightarrow \text{Preço}$
- $\{\text{Vendedor_Id}, \text{Nome}\} \rightarrow \text{Endereço}$

Dependência Funcional: Chave



- Atributo, ou combinação de atributos que identifica univocamente uma linha de uma relação
- ▣ Identificação única: todo atributo não-chave é dependente funcionalmente da chave
- ▣ Não-redundância: nenhum atributo na chave pode ser eliminado sem destruir a propriedade de identificação única

Dependência Funcional: Chave

- A_i, A_j, \dots, A_k é uma chave de uma relação R se e somente se:
 - A_i, A_j, \dots, A_k determina todos os outros atributos de R
 - Nenhum subconjunto de A_i, A_j, \dots, A_k determina funcionalmente os atributos restantes de R
- **Atributo não-primo (ANP)** é um atributo que não está contido em nenhuma chave candidata de R . Todo atributo que pertence a uma chave candidata é denominado primo.
- Uma **superchave** de R é um subconjunto dos atributos de R que contém uma chave.
 - Toda chave é uma superchave, mas nem toda superchave é chave

Dependência Funcional: Chave



- Quando uma relação possui mais de uma chave primária, cada uma é chamada de **chave candidata**.
- Neste caso uma delas é arbitrariamente escolhida para ser a **chave primária**, e as outras são chamadas de **chaves secundárias**.

Dependência Funcional: Chave

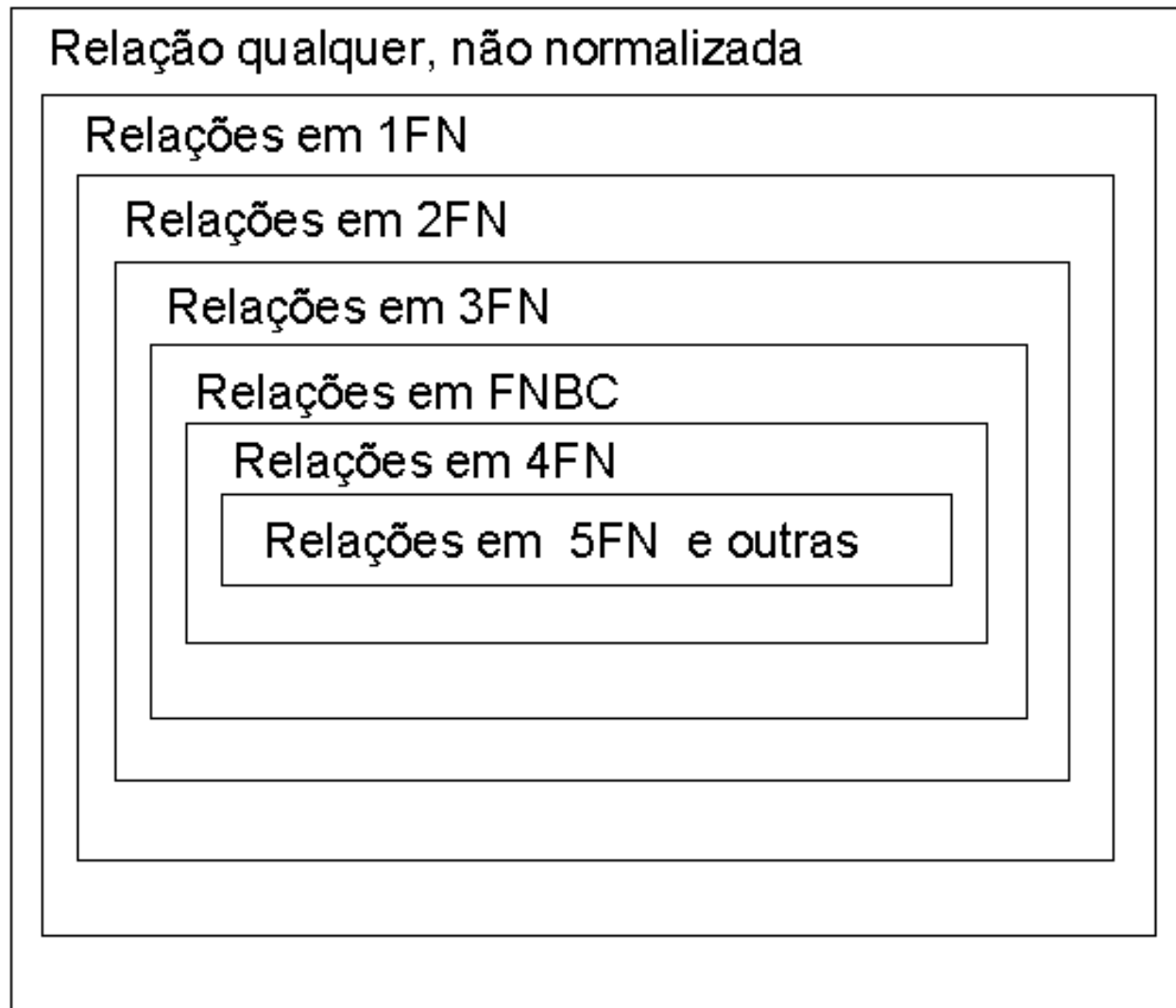
- Um atributo de uma relação R é chamado de **primo de R se for membro de uma chave** candidata de R.
 - Em caso contrário, o atributo é não primo.
 - Ex.:

WORKS_ON (SSN*, PNUMBER*, HOURS)

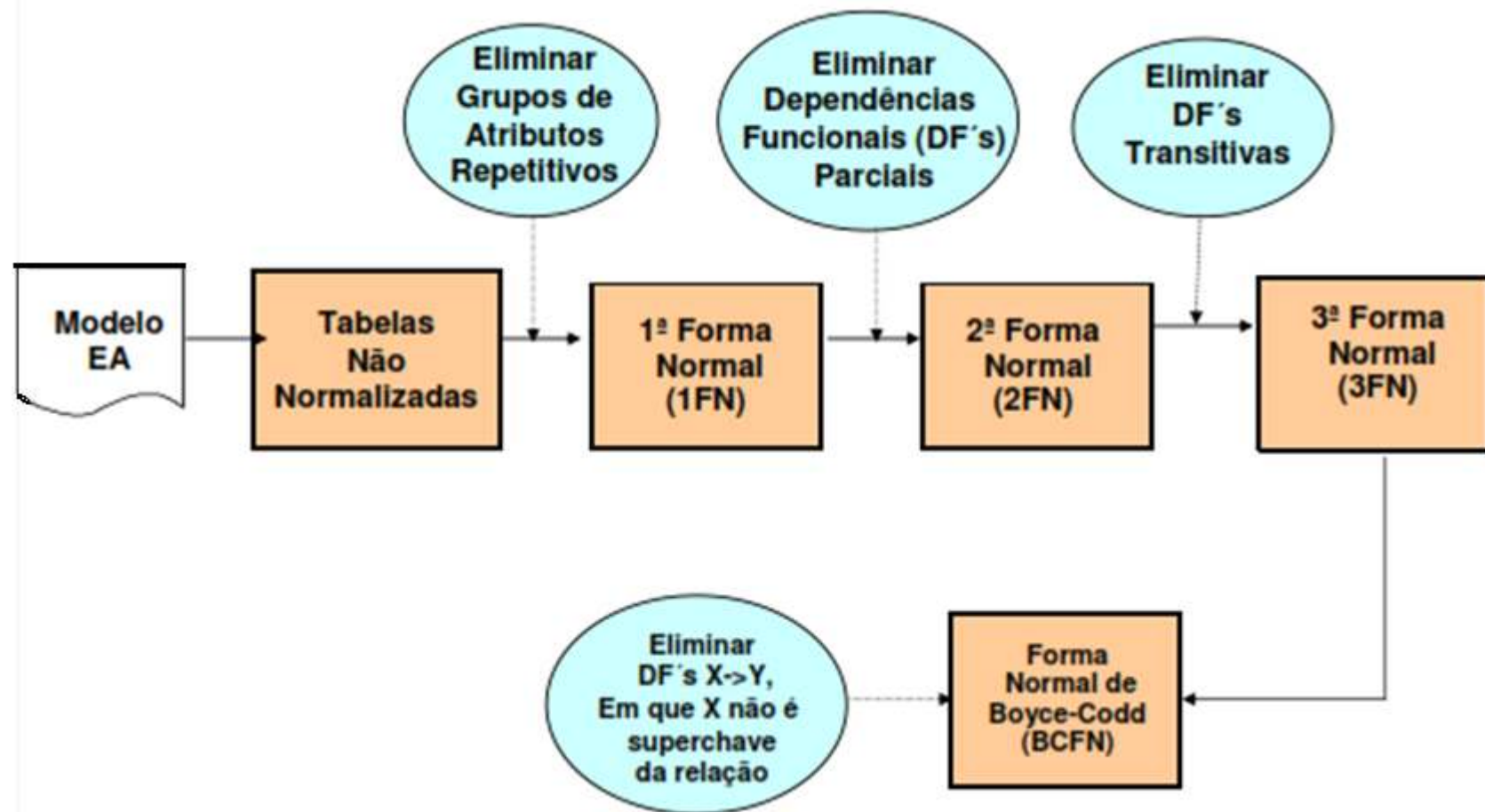
SSN e PNUMBER são primos

HOURS é não primo

Normalização



Etapas



Primeira Forma Normal (1FN)

- É parte da definição formal de uma relação.
- Foi definida para não permitir atributos multivalorados, atributos compostos e suas combinações.

“Uma relação está em 1FN se e somente se todos os seus atributos contêm apenas valores atômicos (simples, indivisíveis).”

Primeira Forma Normal (1FN)

Relação que não está em 1FN

DEPARTAMENTO

Nome	<u>Núm</u>	Gerente	Data-Inicio	Localização
Pesq	5	333445555	22Mai88	Lins, Bauru, Santos
Adm	4	987654321	01Jan92	Campinas
Sede	1	888665555	19Jun91	Santos

atributo multivalorado

Relações em 1FN sem redundância

DEPARTAMENTO

Nome	<u>Núm</u>	Gerente	Data-Inicio
Pesq	5	333445555	22Mai88
Adm	4	987654321	01Jan92
Sede	1	888665555	19Jun91

LOCAL_DEPT

<u>Num-Dept</u>	<u>Localização</u>
1	Santos
4	Campinas
5	Lins
5	Bauru
5	Santos

Dependência Funcional

- Em uma tabela relacional, diz-se que uma coluna C2 depende funcionalmente de uma coluna C1 (ou que uma coluna C1 determina a coluna C2) quando, em todas as linhas da tabela, para cada valor de C1 que aparece na tabela, aparece o mesmo valor de C2.

$$C1 \rightarrow C2$$

- DF deve ser explicitamente definida por alguém que conheça a semântica dos atributos de uma relação.

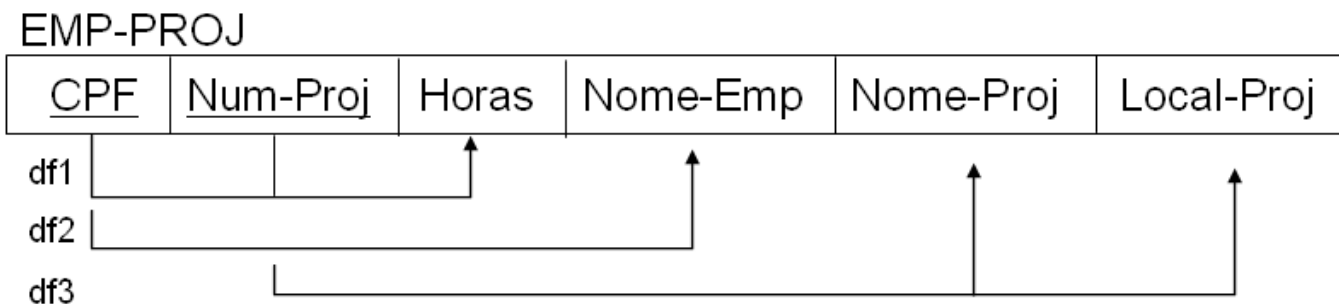
Segunda Forma Normal (2FN)

“Uma relação encontra-se na 2FN se e somente se estiver em 1FN e não contém dependências parciais.”

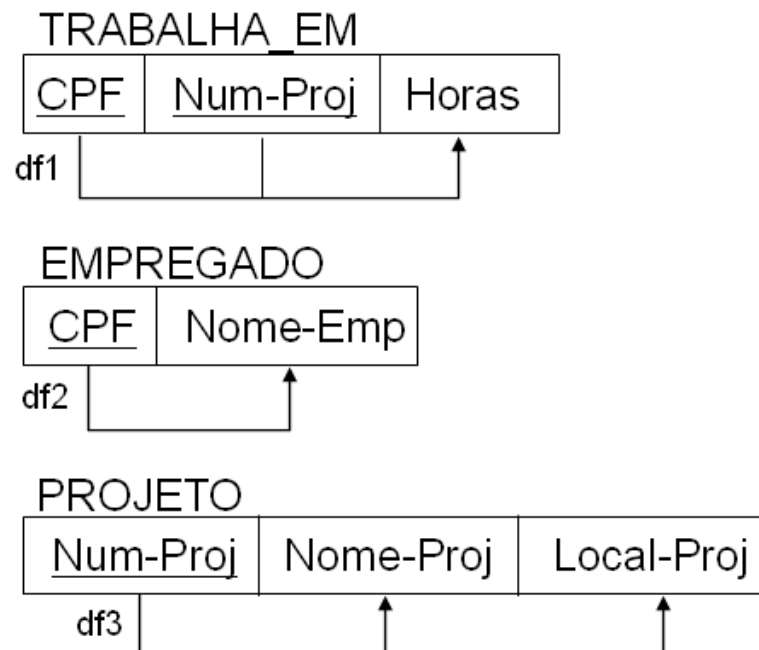
- Dependência Parcial: ocorre quando uma coluna depende apenas de uma parte de uma chave primária composta.

Segunda Forma Normal (2FN)

Relação em 1FN que não está em 2FN



Relações em 2FN



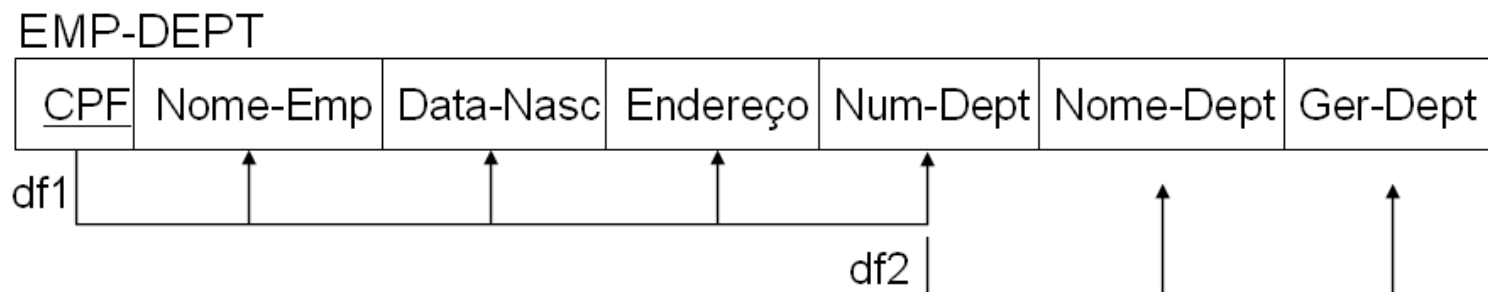
Terceira Forma Normal (3FN)

“Uma relação está em 3FN se e somente se estiver em 2FN e nenhum atributo não-primo (isto é, que não seja membro de uma chave) for transitivamente dependente da chave primária.”

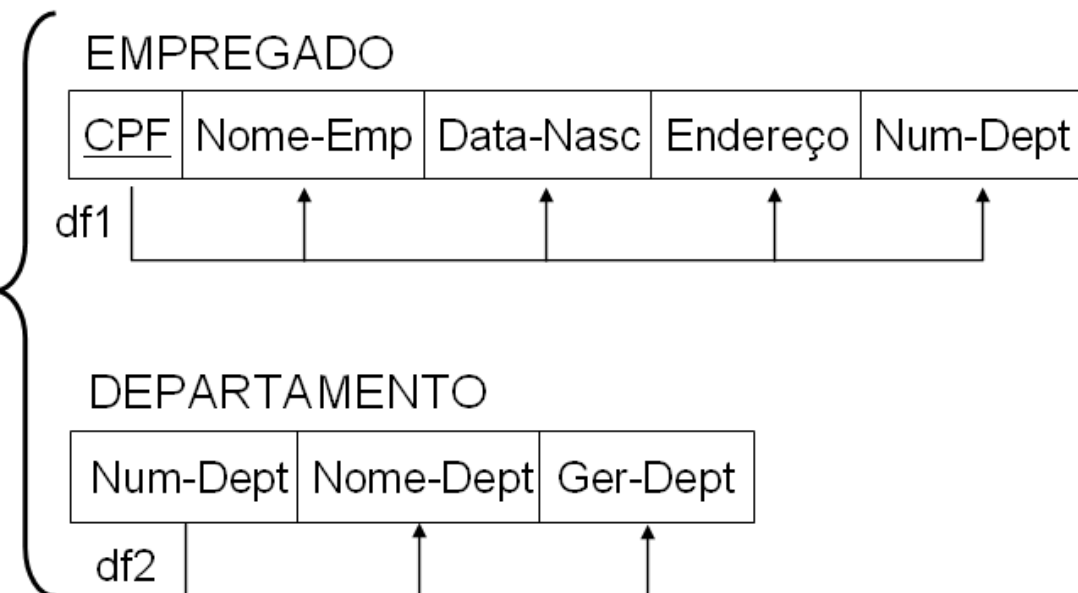
- Dependência Transitiva: ocorre quando uma coluna, além de depender da chave primária de uma tabela, depende de outra coluna ou conjunto de colunas da tabela.
- A relação não deve ter um atributo não-chave funcionalmente determinado por um outro atributo não-chave(ou por um conjunto de atributos não-chave).

Terceira Forma Normal (3FN)

Relação em 2FN que não está em 3FN



Relações em 3FN



Terceira Forma Normal (3FN)

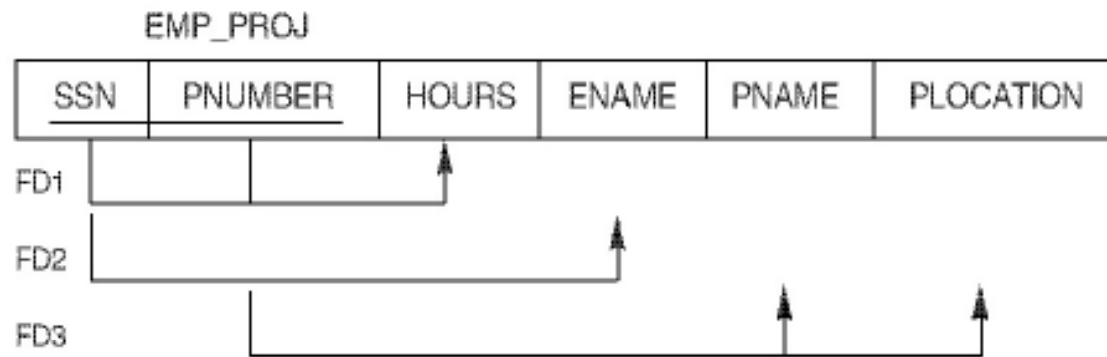
EMP-DEPT

<u>CPF</u>	Nome-Emp	Data-Nasc	Endereço	Num-Dept	Nome-Dept	Ger-Dept
------------	----------	-----------	----------	----------	-----------	----------

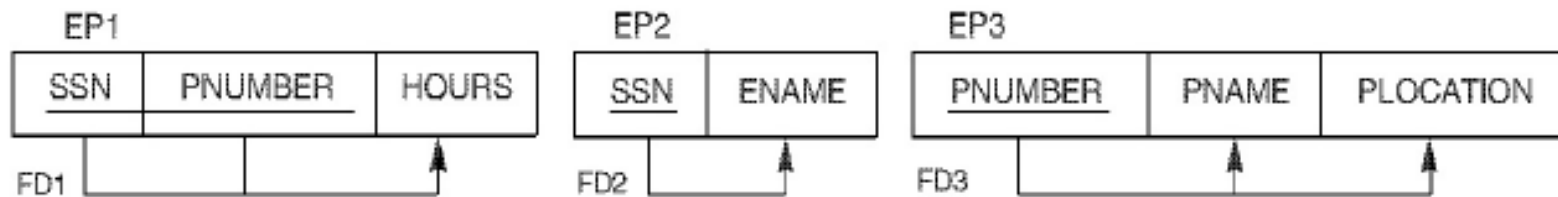
- ❑ Relações não normalizadas são sujeitas a anomalias durante as atualizações:
- ❑ Anomalias de inserção:
 - ❑ Inserir empregado requer repetir dados de departamento.
- ❑ Anomalias de exclusão:
 - ❑ Excluir único empregado de departamento também exclui o departamento.
- ❑ Anomalias de modificação:
 - ❑ Mudar gerente de departamento requer modificar várias tuplas.

Exemplo 3FN

(a)

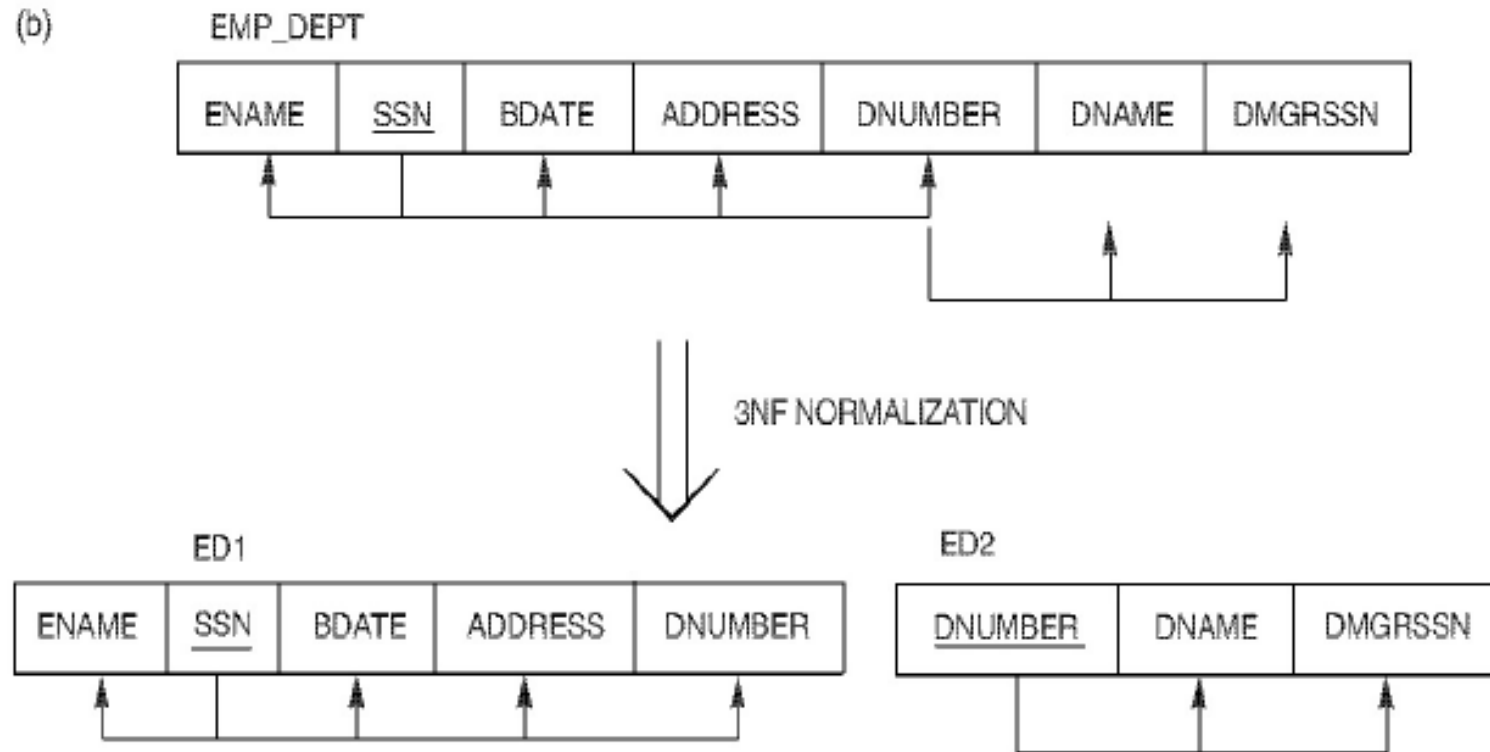


2NF NORMALIZATION



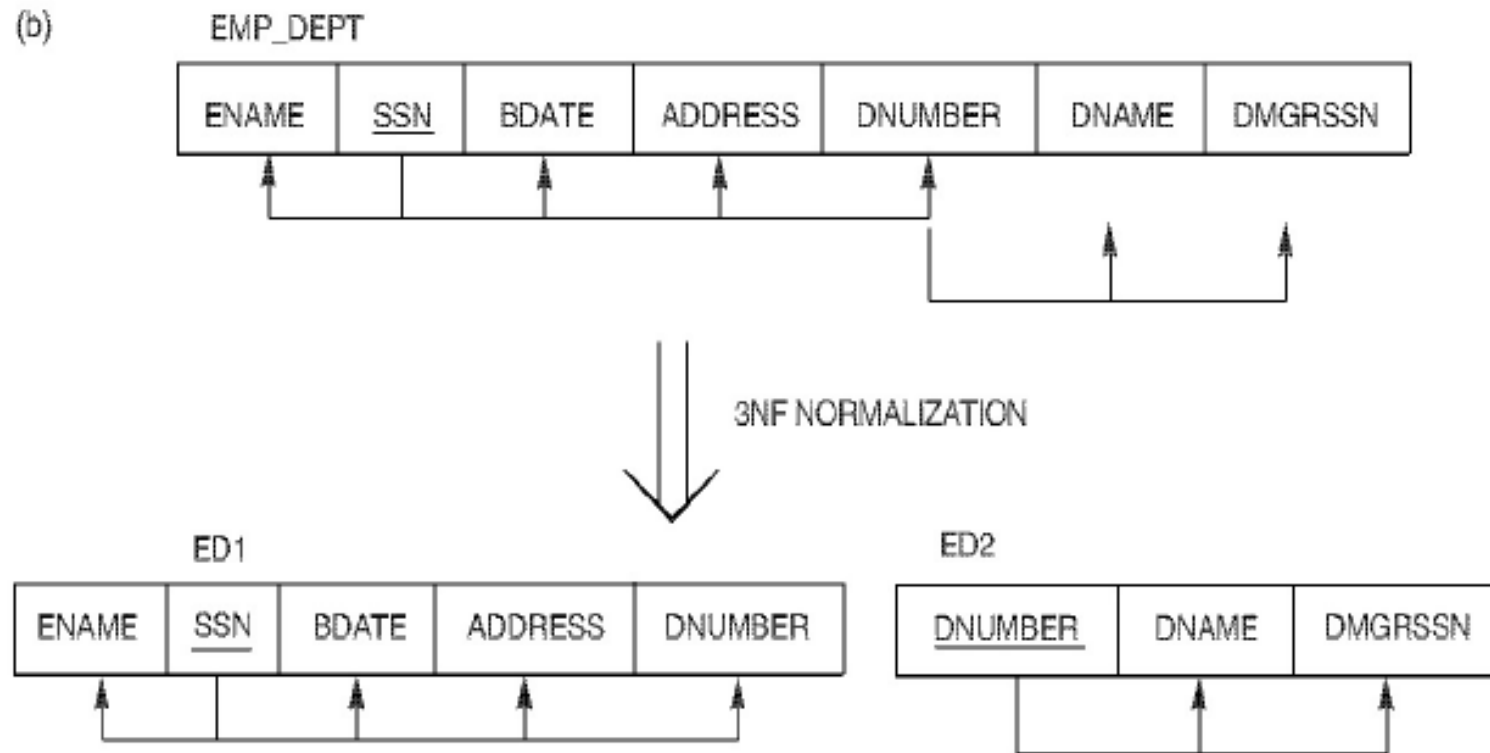
Exemplo 3FN

(b)



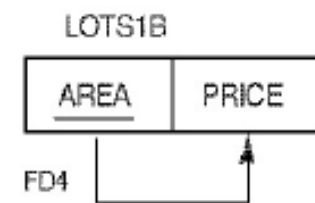
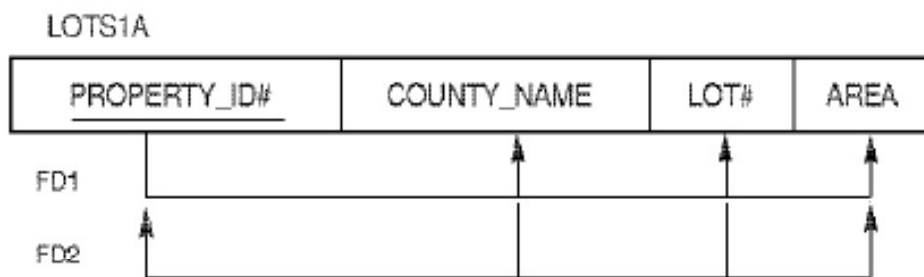
Exemplo2 - 3FN

(b)

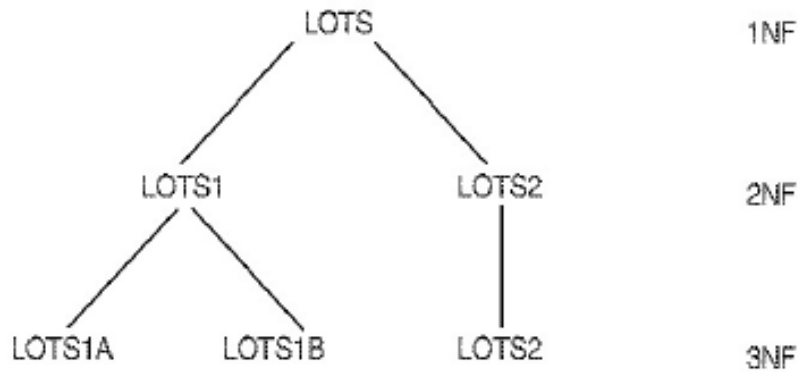


Exemplo2 - 3FN

(c)



(d)



Forma Normal de Boyce Codd (FNBC)

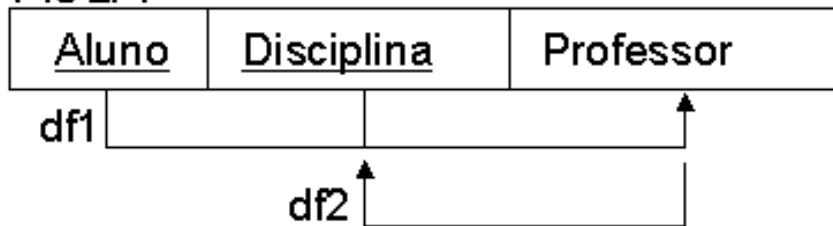
“Uma relação está em FNBC se para toda $df\ X \rightarrow Z$, X é uma super-chave.”

- É uma forma mais restritiva de 3FN, isto é toda relação em FNBC está também em 3FN; entretanto, uma relação em 3FN não está necessariamente em FNBC.
- Uma superchave SK especifica uma restrição de unicidade de que duas tuplas distintas em um estado r de R não podem possuir o mesmo valor para SK.
- Chave primária é a superchave mínima.

Forma Normal de Boyce Codd (FNBC)

Relação em 3FN que não está em FNBC

AULA



df2 : {Professor} → {Disciplina}
{Professor} não é uma superchave.

AULA

<u>Aluno</u>	<u>Disciplina</u>	Professor
Carlos	Química	Ana
Carlos	Física	Antonio
Marta	Química	Ana
Marta	Português	Maria
João	Português	Manoel

Anomalia de exclusão : Se Carlos sair da aula de Física, não teremos nenhum registro de que Antonio leciona Física.

Forma Normal de Boyce Codd (FNBC)

R1

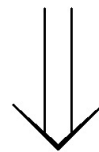
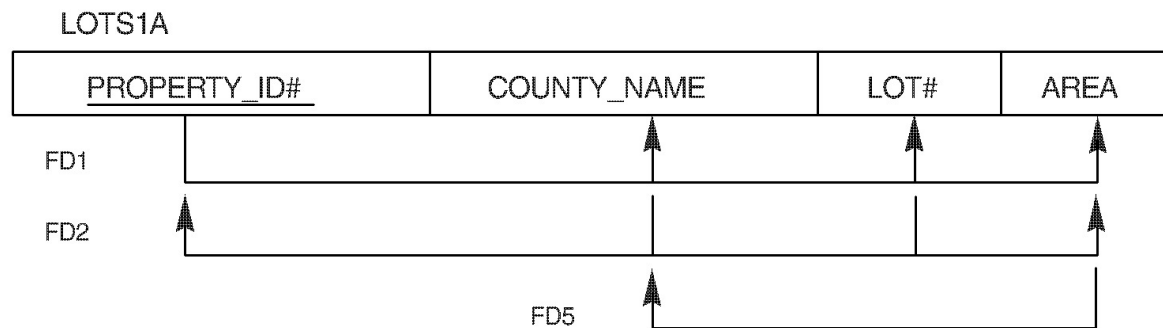
<u>Aluno</u>	<u>Professor</u>
Carlos	Ana
Carlos	Antonio
Marta	Ana
Marta	Maria
João	Manoel

R2

<u>Professor</u>	Disciplina
Ana	Química
Antonio	Física
Maria	Português
Manoel	Português

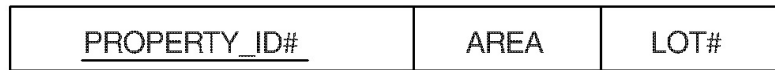
Exemplo FNBC

(a)

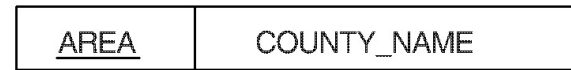


BCNF Normalization

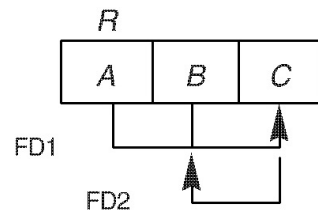
LOTS1AX



LOTS1AY



(b)



Outras Formas Normais

- Normalização é uma Ferramenta para Validação da Qualidade de um Esquema.
- As formas normais até FNBC são baseadas em dependências funcionais, exceto a 1FN, que faz parte da definição do modelo relacional.
- O design conceitual baseado no modelo ER tende naturalmente a produzir esquemas normalizados, a menos da 1FN.
- A separação de conceitos é o resultado natural do design conceitual bem feito.
- Na prática, esquemas que violam a normalização são exemplos de esquemas mal projetados.
- A utilidade prática da 4FN e 5FN é limitada, porque num banco de dados real com muitos atributos, é muito difícil (e praticamente irrelevante) descobrir tais dependências e restrições.

Quarta Forma Normal (4FN)

“Uma relação está em 4ª Forma Normal (4FN) se, e somente se, estiver na 3FN e não contiver dependências multivaloradas.”

- Dada uma relação qualquer com três atributos **x**, **y** e **z**, diz-se que **y** depende de forma multivalorada de **x** se e somente se sempre que existirem duas tuplas (**x1**,**y1**,**z1**) e (**x1**,**y2**,**z2**) existirão também duas tuplas (**x1**,**y1**,**z2**) e (**x1**,**y2**,**z1**).
- Refere-se à combinação de valores de atributos multivalorados disjuntos (**y** e **z**).
- **x** na verdade, relaciona-se com **y** e com **z** de forma independente.

Quarta Forma Normal (4FN)

AutoresAssuntosLivros

No_tombo	Autor	Assunto
.....
321321	Silberschatz, A.	Banco de Dados
321321	Korth, H.F.	Banco de Dados
321321	Silberschatz, A.	SQL
321321	Korth, H.F.	SQL
.....

□ No_tombo $\rightarrow \rightarrow$ Autor

■ No_tombo $\rightarrow \rightarrow$ Assunto

AssuntosLivros

No_tombo	Assunto
.....
321321	Banco de Dados
321321	SQL
.....

AutoresLivros

No_tombo	Autor
.....
321321	Silberschatz, A.
321321	Korth, H.F.
.....

Exemplo 4FN

Eliminação da Multiplicidade de Fatos Multivalorados.

Um paciente pode necessitar de vários exames e dispor de vários planos de saúde.

Portanto, plano de saúde e exame são fatos multivalorados em relação ao paciente.

O paciente João dispõe de dois planos de saúde e necessita fazer dois exames.

Solicitação de Exame

Paciente	Plano de Saúde	Exame
João	Amil	Teor alcoólico
João	Blue-Life	Teor alcoólico
João	Amil	Sangue
João	Blue-Life	Sangue

Exemplo 4FN

Solicitação de Exame

Paciente	Plano de Saúde	Exame
João	Amil	Teor alcoólico
João	Blue-Life	Teor alcoólico
João	Amil	Sangue
João	Blue-Life	Sangue

A tabela está normalizada na 3FN/FNBC (a chave é composta pela união de três atributos). Ainda assim, está difícil manipular mais de um fato multivalorado na mesma tabela.

Resolvemos à situação fazendo a projeção dos subconjuntos que constituem cada fato multivalorado.

Solicitação de Exame

<u>Paciente</u>	<u>Exame</u>
João	Teor alcoólico
João	Sangue

Suporte de Seguridade

<u>Paciente</u>	<u>Plano de Saúde</u>
João	Amil
João	Blue-Life

Quinta Forma Normal (5FN)

Revendas

	Representante	Cliente	Produto
1	Martins	Sá e filhos	Louça
3	Martins	DMI, Lda	Louça
4	Martins	DMI, Lda	Talheres
5	Castro	Sá e filhos	Cintos
7	Castro	Sá e filhos	Lenços
9	Castro	YSL	Meias
10	Castro	YSL	Lenços

Clientes_Representante

Representante	Cliente
Martins	Sá e filhos
Martins	DMI, Lda
Castro	Sá e filhos
Castro	YSL

Representação

Representante	Produto
Martins	Louça
Martins	Talheres
Castro	Cintos
Castro	Meias
Castro	Lenços

Quinta Forma Normal (5FN)

Revendas

	Representante	Cliente	Produto
1	Martins	Sá e filhos	Louça
3	Martins	DMI, Lda	Louça
4	Martins	DMI, Lda	Talheres
5	Castro	Sá e filhos	Cintos
7	Castro	Sá e filhos	Lenços
9	Castro	YSL	Meias
10	Castro	YSL	Lenços

Clientes_Representante

Represen.	Cliente
Martins	Sá e filhos
Martins	DMI, Lda
Castro	Sá e filhos
Castro	YSL

Representação

Represen.	Produto
Martins	Louça
Martins	Talheres
Castro	Cintos
Castro	Meias
Castro	Lenços

Vendas_Clientes

Cliente	Produto
Sá e filhos	Louça
Sá e filhos	Cintos
Sá e filhos	Lenços
DMI, Lda	Louça
DMI, Lda	Talheres
YSL	Meias
YSL	Lenços

Quinta Forma Normal (5FN)

- Existem relações que não podem ser decompostas em duas projeções sem perda, mas podem ser decompostas em três ou mais. Estas relações podem ser descritas como “decomponível n ” ($n > 2$) (Date), significando que a relação em questão pode ser decomposta sem perda em n projeções, mas não em m projeções, $m < n$.
- Esta limitação é denominada *dependência de junção* (DJ).
- Uma relação R satisfaz a dependência de junção * $(X, Y, \dots Z)$ se, e somente se, R for igual à junção de suas projeções em $X, Y, \dots Z$, onde $X, Y, \dots Z$ forem subconjuntos do conjunto de atributos de R .

Quinta Forma Normal (5FN)

“Uma relação R está na 5FN , também chamada de forma normalizada de projeção-junção (PJ/NF) se, e somente se, toda dependência de junção em R for consequência de chaves candidatas de R .”

- As DM são uma tentativa de detectar decomposições sem perdas que se apliquem a todas as relações de um dado esquema.
- Se não é possível reconhecer qualquer DM em R , não existe decomposição sem perdas em duas relações.